

Amendments To The Claims:

This listing of claims will replace all prior versions and listings of claims in the application:

Claims 1-111. (cancelled)

Claim 112. (currently amended) A method comprising:

- (a) providing first data from a first set of samples wherein:
 - (i) the first set of samples comprises a plurality of samples classified into a first biological state class and a plurality of samples classified into a second biological state class;
 - (ii) the data from ~~each sample in the first set~~ of samples comprises a plurality of data elements, each data element characterized by a value, wherein all of the samples share a plurality of common data elements;
- (b) performing multivariate analysis on the first data to qualify each common data element in the first data based on the ability of the data element to classify a sample into the first biological state class or the second biological state class, wherein classification is as a function of data element value;
- (c) selecting a first subset of qualified common data elements from the first data;
- (d) providing second data from a second set of samples wherein:
 - (i) the second set of samples comprises a plurality of samples classified into the first biological state class and a plurality of samples classified into the second biological state class;
 - (ii) the data from ~~each sample in the second set~~ of samples comprises a plurality of data elements, each data element characterized by a value, wherein all of the samples share the plurality of common data elements;
 - (iii) the first samples and second samples come from first and second populations that have a statistically significant difference with respect to at least one preanalytical variable;

- (e) performing multivariate analysis on the second data to qualify each common data element in the second data based on the ability of the data element to classify a sample into the first biological state class or the second biological state class, wherein classification is as a function of data element value;
- (f) selecting a second subset of qualified common data elements from the second data;
- (g) selecting an intersection subset of data elements from the first and second subsets, wherein each data element in the intersection subset is a member of both of the first and second subsets; and
- (h) displaying the intersection subset on a graphical display interface on a user device.

Claim 113. (previously presented) The method of claim 112 wherein the first and second populations have a statistically significant difference with respect to a preanalytical variable selected from the group consisting of gender, age, ethnicity, sample collection parameter, sample processing parameter, weight, diet, medication status, medical condition, amount of physical exercise, pregnancy, level of circulating antibodies and a clinical characteristic.

Claim 114. (previously presented) The method of claim 113 wherein the first and second populations have a statistically significant difference with respect to a plurality of preanalytical variables selected from said group.

Claims 115. (previously presented) The method of claim 112 wherein the first samples and the second samples are collected from different geographical locations.

Claims 116. (previously presented) The method of claim 112 wherein the first samples and the second samples are collected from different clinical trial sites.

Claim 117. (previously presented) The method of claim 112 wherein the step of selecting the first and second subsets comprises using the discovery data sets to train a learning

algorithm wherein the learning algorithm ranks the data elements based on a quantitative measure of ability to classify.

Claim 118. (previously presented) The method of claim 117 wherein the learning algorithm is a supervised learning algorithm.

Claim 119. (previously presented) The method of claim 117 wherein the learning algorithm is an unsupervised learning algorithm.

Claim 120. (previously presented) The method of claim 117 wherein the training comprises using support vector machine analysis.

Claim 121. (previously presented) The method of claim 117 wherein the training comprises performing linear discrimination analysis.

Claim 122. (previously presented) The method of claim 117 wherein the training comprises performing unified maximum separability analysis (UMSA).

Claim 123. (previously presented) The method of claim 112 further comprising independently re-sampling data elements in each data set.

Claim 124. (previously presented) The method of claim 112 further comprising, selecting candidate biomarkers from selected data elements and testing one or more of the candidate biomarkers on a validation data set.

Claim 125. (previously presented) The method of claim 112 wherein the biological state class comprises a cell state.

Claim 126. (previously presented) The method of claim 112 wherein the biological state class is a patient status.

Claim 127. (previously presented) The method of claim 112 wherein the biological state class is selected from the group consisting of: presence of a disease; absence of a disease; progression of a disease; risk for a disease; stage of disease; likelihood of recurrence of disease; a genotype; a phenotype; exposure to an agent or condition; a demographic characteristic; resistance to agent, sensitivity to an agent, and combinations thereof.

Claim 128. (previously presented) The method of claim 127 wherein the genotype is selected from the group consisting of an HLA haplotype; a mutation in a gene; a modification of a gene, and combinations thereof.

Claim 129. (previously presented) The method of claim 127 wherein the agent is selected from the group consisting of a toxic substance, a potentially toxic substance, an environmental pollutant, a candidate drug, and a known drug.

Claim 130. (previously presented) The method of claim 127 wherein sensitivity to an agent comprises responsiveness to a drug.

Claim 131. (previously presented) The method of claim 124 wherein the one or more candidate biomarkers are diagnostic of the presence of a disease, risk of developing a disease, risk of recurrence of a disease, or stage of the disease.

Claim 132. (previously presented) The method of claim 112 wherein values of the data elements in a data point represent levels and/or frequency of components in a data point sample.

Claim 133. (previously presented) The method of claim 132 wherein components are selected from the group consisting of: nucleic acids, proteins, polypeptides, peptides, carbohydrates and modified or processed forms thereof.

Claim 134. (previously presented) The method of claim 112 wherein levels of components are measured by an expression profiling assay.

Claim 135. (previously presented) The method of claim 134 wherein the expression profiling assay comprises measuring the amount and/or form of a nucleic acid.

Claim 136. (previously presented) The method of claim 134 wherein expression profiling comprises measuring amplification, mutation, and/or modification of DNA.

Claim 137. (previously presented) The method of claim 134 wherein the expression profiling assay comprises measuring the amount and/or form of a protein, polypeptide or peptide.

Claim 138. (previously presented) The method of claim 137 wherein the expression profiling assay comprises mass spectrometry.

Claim 139. (previously presented) The method of claim 138 wherein the expression profiling assay comprises SELDI analysis.

Claim 140. (previously presented) The method of claim 134 wherein the expression profiling assay comprises measuring the amount and/or form of a carbohydrate.

Claim 141. (previously presented) The method of claim 112 wherein expression profiling comprises:

- (a) contacting samples with a substrate comprising binding partners for specifically binding to sample components having selected characteristics and
- (b) identifying sample components bound to the substrate.

Claim 142. (previously presented) The method of claim 141 wherein binding partners are selected from the group consisting of cationic molecules; anionic molecules; metal chelates; antibodies; single- or double-stranded nucleic acids; proteins, peptides, amino acids; carbohydrates; lipopolysaccharides; sugar amino acid hybrids; molecules from phage display libraries; biotin; avidin; streptavidin; and combinations thereof.

Claim 143. (previously presented) The method of claim 141 wherein the binding partners are arrayed on the substrate.

Claim 144. (previously presented) The method of claim 117 wherein an assay used to measure levels of data elements in training data sets from which candidate biomarkers are identified is different from an assay used to measure data elements in a validation data set used to validate the candidate biomarker.

Claim 145. (previously presented) The method of claim 140 wherein the assay used to measure levels of data elements in training data sets is SELDI.

Claim 146. (previously presented) The method of claim 140 wherein the assay used to measure levels of data elements in validation data sets is an immunoassay.

Claim 147. (previously presented) The method of claim 112 wherein the independent discovery data sets are collected from different locations, using different collection protocols, and/or are collected from different populations.

Claim 148. (previously presented) The method of claim 112 wherein each discovery data set is from a different clinical trial site.

Claim 149. (currently amended) A computer program product comprising a written, electronic, magnetic or optical physical media that is computer readable and having—comprising a computer readable medium having:

- (a) receiving input data ~~of relating to~~ at least first and second independent discovery data sets wherein:
 - (i) the data sets comprise a plurality of forms of biological state classes;
 - (ii) each data set comprises a plurality of data points, wherein each data point exhibits one form of a biological state class and each data set

comprises a plurality of data points belonging to each of the classes;
and

(iii) each data point comprises a plurality of data elements, each data element characterized by a value, wherein all data points share a plurality of common data elements;

(b) a second computer readable program code providing instructions for qualifying each common data element, independently for each data set, based on the ability of the data element to classify a data point into a biological state class, as a function of data element value and for selecting an initial subset of data elements within each data set, and

(c) a third computer readable program code providing instructions for selecting an intersection subset of data elements from the initial subsets, wherein each data element in the intersection subset is a member of a majority of the initial subsets.

Claim 150. (previously presented) The computer program product of claim 149 wherein selecting the initial subsets comprises using the discovery data sets to train a learning algorithm wherein the learning algorithm ranks the data elements based on a quantitative measure of ability to classify.

Claim 151. (previously presented) The computer program product of claim 149 wherein the learning algorithm is a supervised learning algorithm.

Claim 152. (previously presented) The computer program product of claim 149 wherein the learning algorithm is an unsupervised learning algorithm.

Claim 153. (previously presented) The computer program product of claim 150 wherein training comprises support vector machine analysis.

Claim 154. (previously presented) The computer program product of claim 150 wherein training comprises linear discrimination analysis.

Claim 155. (previously presented) The computer program product of claim 150 wherein training comprises combining support vector machine analysis and linear discrimination analysis.

Claim 156. (previously presented) The computer program product of claim 150 wherein training comprises performing unified maximum separability analysis (UMSA).

Claim 157. (previously presented) The computer program product of claim 149 further comprising program code for independently re-sampling data elements in each data set.

Claim 158. (previously presented) The computer program product of claim 149 further comprising program code for selecting candidate biomarkers based on ranking by the learning algorithm and for testing one or more of the candidate biomarkers on a validation data set.

Claim 159. (previously presented) The computer program product of claim 149 wherein the biological state class comprises a cell state.

Claim 160. (previously presented) The computer program product of claim 149 wherein the biological state class comprises a patient status.

Claim 161. (previously presented) The computer program product of claim 149 wherein the biological state class is selected from the group consisting of: presence of a disease; absence of a disease; progression of a disease; risk for a disease; stage of disease; likelihood of recurrence of disease; a genotype; a phenotype; exposure to an agent or condition; a demographic characteristic; resistance to agent, sensitivity to an agent, and combinations thereof.

Claim 162. (previously presented) The computer program product of claim 161 wherein the genotype is selected from the group consisting of an HLA haplotype; a mutation in a gene; a modification of a gene, and combinations thereof.

Claim 163. (previously presented) The computer program product of claim 161 wherein the agent is selected from the group consisting of a toxic substance, a potentially toxic substance, an environmental pollutant, a candidate drug, and a known drug.

Claim 164. (previously presented) The computer program product of claim 161 wherein sensitivity to an agent comprises responsiveness to a drug.

Claim 165. (previously presented) The computer program product of claim 158 wherein the one or more candidate biomarkers are diagnostic of the presence of a disease, risk of developing a disease, risk of recurrence of a disease, or stage of the disease.

Claim 166. (previously presented) The computer program product of claim 161 wherein values of the data elements in a data point represent levels and/or frequency of components in a data point sample.

Claim 167. (previously presented) The computer program product of claim 161 wherein components are selected from the group consisting of: nucleic acids, proteins, polypeptides, peptides, carbohydrates and modified or processed forms thereof.

Claim 168. (previously presented) The computer program product of claim 160 wherein levels of components are measured by an expression profiling assay.

Claim 169. (previously presented) The computer program product of claim 168 wherein the expression profiling assay comprises measuring the amount and/or form of a nucleic acid.

Claim 170. (previously presented) The computer program product of claim 168 wherein expression profiling comprises measuring amplification, mutation, and/or modification of DNA.

Claim 171. (previously presented) The computer program product of claim 168 wherein the expression profiling assay comprises measuring the amount and/or form of a protein, polypeptide or peptide.

Claim 172. (previously presented) The computer program product of claim 168 wherein the expression profiling assay comprises mass spectrometry.

Claim 173. (previously presented) The computer program product of claim 168 wherein the expression profiling assay comprises SELDI analysis.

Claim 174. (previously presented) The computer program product of claim 168 wherein the expression profiling assay comprises measuring the amount and/or form of a carbohydrate.

Claim 175. (previously presented) The computer program product of claim 168 wherein expression profiling comprises:

- (a) contacting samples with a substrate comprising binding partners for specifically binding to sample components having selected characteristics; and
- (b) identifying sample components bound to the substrate.

Claim 176. (previously presented) The computer program product of claim 175 wherein binding partners are selected from the group consisting of cationic molecules; anionic molecules; metal chelates; antibodies; single- or double-stranded nucleic acids; proteins, peptides, amino acids; carbohydrates; lipopolysaccharides; sugar amino acid hybrids; molecules from phage display libraries; biotin; avidin; streptavidin; and combinations thereof.

Claim 177. (previously presented) The computer program product of claim 149 wherein an assay used to measure levels of data elements in training data sets from which candidate biomarkers are identified is different from an assay used to measure data elements in a validation data set used to validate the candidate biomarker.

Claim 178. (previously presented) The computer program product of claim 149 wherein the assay used to measure levels of data elements in training data sets is SELDI.

Claim 179. (previously presented) The computer program product of claim 149 wherein the assay used to measure levels of data elements in validation data sets is an immunoassay.

Claim 180. (previously presented) The computer program product of claim 149 wherein the independent discovery data sets are collected from different locations, using different collection protocols, and/or are collected from different populations.

Claim 181. (previously presented) The computer program product of claim 149 wherein each discovery data set is from a different clinical trial site.

Claim 182. (currently amended) A system comprising:

one or more processors for

(a) receiving input data comprising ~~relating to~~ at least first and second independent discovery data sets wherein:

(i) the first set of samples comprises a plurality of samples classified into a first biological state class and a plurality of samples classified into a second biological state class;

(ii) the data from ~~each sample in the first~~ sample set comprises a plurality of data elements, each data element characterized by a value, wherein all of the samples share a plurality of common data elements;

(iii) the second set of samples comprises a plurality of samples classified into the first biological state class and a plurality of samples classified into the second biological state class;

(iv) the data from ~~each sample in the second~~ sample set comprises a plurality of data elements, each data element characterized by a value, wherein all of the samples share the plurality of common data elements;

(b) executing computer readable program code providing instructions for qualifying each common data element, independently for each data set, based on the ability of the data element to classify a data point into a biological state class, wherein classification is as a function of data element value and for selecting an initial subset of data elements within each data set; and

(c) executing computer readable program code providing instructions for selecting an intersection subset of data elements from the initial subsets, wherein each data element in the intersection subset is a member of a majority of the initial subsets.

Claim 183. (previously presented) The system of claim 182 further comprising one or more devices for providing input data to the one or more processors.

Claim 184. (previously presented) The system of claim 182 wherein the one or more devices for providing input data comprises a detector for detecting a characteristic of a data element.

Claim 185. (previously presented) The system of claim 182 wherein the detector comprises a mass spectrometer.

Claim 186. (previously presented) The system of claim 182 wherein the detector comprises a gene chip reader.

Claim 187. (previously presented) The system of claim 182 further comprising a memory for storing a data set of ranked data elements.

Claim 188. (previously presented) The system of claim 182 further comprising a database of ranked data elements.

Claim 189. (previously presented) The system of claim 182 wherein selecting the initial subsets comprises using the discovery data sets to train a learning algorithm wherein

the learning algorithm ranks the data elements based on a quantitative measure of ability to classify.

Claim 190. (previously presented) The system of claim 189 wherein the learning algorithm is a supervised learning algorithm.

Claim 191. (previously presented) The system of claim 189 wherein the learning algorithm is an unsupervised learning algorithm.

Claim 192. (previously presented) The system of claim 189 wherein training comprises support vector machine analysis.

Claim 193. (previously presented) The system of claim 189 wherein training comprises linear discrimination analysis.

Claim 194. (previously presented) The system of claim 189 wherein training comprises combining support vector machine analysis and linear discrimination analysis.

Claim 195. (previously presented) The system of claim 189 wherein training comprises performing unified maximum separability analysis (UMSA).

Claim 196. (previously presented) The system of claim 182 wherein the system further executes program code for independently re-sampling data elements in each data set.

Claim 197. (previously presented) The system of claim 189 wherein the system further executes program code for selecting candidate biomarkers based on ranking by the learning algorithm and for testing one or more of the candidate biomarkers on a validation data set.

Claim 198. (previously presented) The system of claim 182 wherein the biological state class comprises a cell state.

Claim 199. (previously presented) The system of claim 182 wherein the biological state class comprises a patient status.

Claim 200. (previously presented) The system of claim 182 wherein the biological state class is selected from the group consisting of: presence of a disease; absence of a disease; progression of a disease; risk for a disease; stage of disease; likelihood of recurrence of disease; a genotype; a phenotype; exposure to an agent or condition; a demographic characteristic; resistance to agent, sensitivity to an agent, and combinations thereof.

Claim 201. (previously presented) The system of claim 200 wherein the genotype is selected from the group consisting of an HLA haplotype; a mutation in a gene; a modification of a gene, and combinations thereof.

Claim 202. (previously presented) The system of claim 200 wherein the agent is selected from the group consisting of a toxic substance, a potentially toxic substance, an environmental pollutant, a candidate drug, and a known drug.

Claim 203. (previously presented) The system of claim 200 wherein sensitivity to an agent comprises responsiveness to a drug.

Claim 204. (previously presented) The system of claim 197 wherein the one or more candidate biomarkers are diagnostic of the presence of a disease, risk of developing a disease, risk of recurrence of a disease, or stage of the disease.

Claim 205. (previously presented) The system of claim 182 wherein values of the data elements in a data point represent levels and/or frequency of components in a data point sample.

Claim 206. (previously presented) The system of claim 205 wherein components are selected from the group consisting of: nucleic acids, proteins, polypeptides, peptides, carbohydrates and modified or processed forms thereof.

Claim 207. (previously presented) The system of claim 205 wherein levels of components are measured by an expression profiling assay.

Claim 208. (previously presented) The system of claim 207 wherein the expression profiling assay comprises measuring the amount and/or form of a nucleic acid.

Claim 209. (previously presented) The system of claim 207 wherein expression profiling comprises measuring amplification, mutation, and/or modification of DNA.

Claim 210. (previously presented) The system of claim 207 wherein the expression profiling assay comprises measuring the amount and/or form of a protein, polypeptide or peptide.

Claim 211. (previously presented) The system of claim 207 wherein the expression profiling assay comprises mass spectrometry.

Claim 212. (previously presented) The system of claim 214 wherein the expression profiling assay comprises SELDI analysis.

Claim 213. (previously presented) The system of claim 207 wherein the expression profiling assay comprises measuring the amount and/or form of a carbohydrate.

Claim 214. (previously presented) The system of claim 207 wherein expression profiling comprises:

- (a) contacting samples with a substrate comprising binding partners for specifically binding to sample components having selected characteristics and
- (b) identifying sample components bound to the substrate.

Claim 215. (previously presented) The system of claim 214 wherein binding partners are selected from the group consisting of cationic molecules; anionic molecules; metal chelates; antibodies; single- or double-stranded nucleic acids; proteins, peptides, amino acids; carbohydrates; lipopolysaccharides; sugar amino acid hybrids; molecules from phage display libraries; biotin; avidin; streptavidin; and combinations thereof.

Claim 216. (previously presented) The system of claim 182 wherein an assay used to measure levels of data elements in training data sets from which candidate biomarkers are identified is different from an assay used to measure data elements in a validation data set used to validate the candidate biomarker.

Claim 217. (previously presented) The system of claim 216 wherein the assay used to measure levels of data elements in training data sets is SELDI.

Claim 218. (previously presented) The system of claim 216 wherein the assay used to measure levels of data elements in validation data sets is an immunoassay.

Claim 219. (previously presented) The system of claim 182 wherein the independent discovery data sets are collected from different locations, using different collection protocols, and/or are collected from different populations.

Claim 220. (previously presented) The system of claim 182 wherein each discovery data set is from a different clinical trial site.

Claim 221. (new) The method of claim 112 wherein the multivariate analysis on the first data comprises use of a pattern recognition process.

Claim 222. (new) The method of claim 221 wherein the pattern recognition process comprises use of a classification model.

Claim 223. (new) The method of claim 112 wherein the multivariate analysis on the second data comprises use of a pattern recognition process.

Claims 224. (new) The method of claim 223 wherein the pattern recognition process comprises use of a classification model.